Final Technical Report

USGS/NEHRP Award No: 08HQGR0043

UPDATING AND MAINTAINING A DOUBLE-DIFFERENCE EARTHQUAKE CATALOG FOR NORTHERN CALIFORNIA

Principal Investigator: Felix Waldhauser

Co-PI: David Schaff

Lamont-Doherty Earth Observatory of Columbia University
Palisades, NY 10964
Tel: (845) 365 8538; Fax: (845) 365 8150
felixw@ldeo.columbia.edu

Project Period: 01/01/2008 – 12/31/2008 (NFE until 12/31/2009)

NEHRP Element III
Key Words: Seismology, Source Characteristics, Database

Research supported by the U.S. Geological Survey (USGS), Department of the Interior, under USGS award number 08HQGR0043. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Government.

Final Technical Report

USGS/NEHRP Award No: 08HQGR0043

UPDATING AND MAINTAINING A DOUBLE-DIFFERENCE EARTHQUAKE CATALOG FOR NORTHERN CALIFORNIA

Principal Investigator: Felix Waldhauser

Co-PI: David Schaff

Lamont-Doherty Earth Observatory of Columbia University
Palisades, NY 10964
Tel: (845) 365 8538; Fax: (845) 365 8150
felixw@ldeo.columbia.edu

ABSTRACT

We have developed, tested, and implemented a real-time procedure (DD-RT) that uses cross-correlation and double-difference methods to rapidly relocate new earthquakes in Northern California relative to nearby reference events found in a high-resolution double-difference (DD) earthquake catalog. The DD-RT software currently runs on a test-bed at the Lamont-Doherty Earth Observatory (LDEO), using near real-time parametric and waveform data feeds from the Northern California Earthquake Data Center (NCEDC) for new events, and a locally stored archive of seismic data for past events. We evaluated the performance of the new monitoring system by back-testing it with events that occurred in the past. The results indicate that the real-time solutions are on average within 0.08 km laterally and 0.24 km vertically of the double-difference catalog locations. The DD-RT system updates the currently employed DD base catalog, which includes events between 1984-2003, with the most recent events in Northern California. Work towards updating the 2003 DD base catalog in order to close the time gap will be completed in the next months. Real-time double-difference location allows for monitoring spatio-temporal changes in seismogenic properties of active faults with unprecedented resolution and therefore has considerable social and economic impact in the immediate evaluation and mitigation of seismic hazards. This study also shows that the precision with which new events are located using this technique will improve with time, helped by the continued increase in density of recorded earthquakes and growth of the digital seismic archives.

1. Overview of Investigations

This final report covers the activities performed between January 1, 2008 (start date of the project) and December 31, 2009. The initial project was for one year, but has been extended by an additional year at no cost to the end of 2009 to accommodate delays in getting the waveform data necessary to carry out the proposed work. The work described in this report is being undertaken by the principle investigator Felix Waldhauser and by co-PI David Schaff.

We originally proposed to update an existing double-difference earthquake catalog for Northern California that includes events between 1984-2003 (Waldhauser and Schaff, 2008) to include events up to the end of 2007. We had to divert from this original plan because of a delay in acquiring the waveform data from the Northern California Earthquake Data Center (NCEDC) for events between 2003-2007. Therefore, in 2008, we completed a prototype system for real-time double-difference event relocation (DD-RT; Waldhauser, 2009). The DD-RT system has been implemented and tested on a Lamont server and automatically updates, using near-real-time data feeds from the USGS and NCEDC, the 2003 DD catalog with double-difference locations of new events. The high-precision hypocenter locations for new events are being posted to www.ldeo.columbia.edu/~felixw/DDrtCISN, typically within minutes after receiving the data.

In Summer 2009, following the acquisition of the waveform data from the NCEDC for events between May 2003 (end of the existing DD catalog) and 2008, we carried out the cross-correlation measurements necessary for relocation (0.5 million delay times in addition to the 3 million that are in the existing data base). A preliminary 2008 edition of the Northern California DD earthquake catalog, including 456,000 events between 1984-2008, has been computed, and we are in the process of testing the robustness of and compute uncertainties for these locations in order to prepare a release of the new catalog. The procedures used to generate the updated DD catalog for Northern CA are, apart from some improvements, similar to the ones described in Waldhauser and Schaff (2008).

2. Investigations undertaken

2.1 A prototype system for real-time double-difference event relocation (DD-RT)

We developed, tested, and implemented a new real-time double-difference location system (DD-RT) on a test-bed here at Lamont, using near-real time data feeds from earthquakes recorded at stations of the Northern California Seismic System (NCSS). The NCSS, which assimilates data from 13 seismic networks (among them NC, NN, WR, CI, PG & UW), records an average of ~50 earthquakes on 1200 channels each day. The core of the DD-RT software is a real-time version of the hypoDD software package (Waldhauser, 2001), with numerous additional programs for collecting, selecting, and processing the data, controlling the process flow, and generating a dynamically updated web site for displaying and accessing the relocation results over the internet (Figure 1). We are running the DD-RT process since 2008 here at Lamont, using the 2003 DD earthquake catalog of Waldhauser and Schaff (2008) as base catalog. Since 2008 we have continuously updated the 2003 DD catalog and released results via our dedicated website www.ldeo.columbia.edu/~felixw/DDrtCISN. We are currently updating the data archives and catalogs (see below) to close the 2003-2009 time gap. In the following we briefly describe the basics of the DD-RT procedure and show results from from operation. This work is described in more detail in Waldhauser (2009).

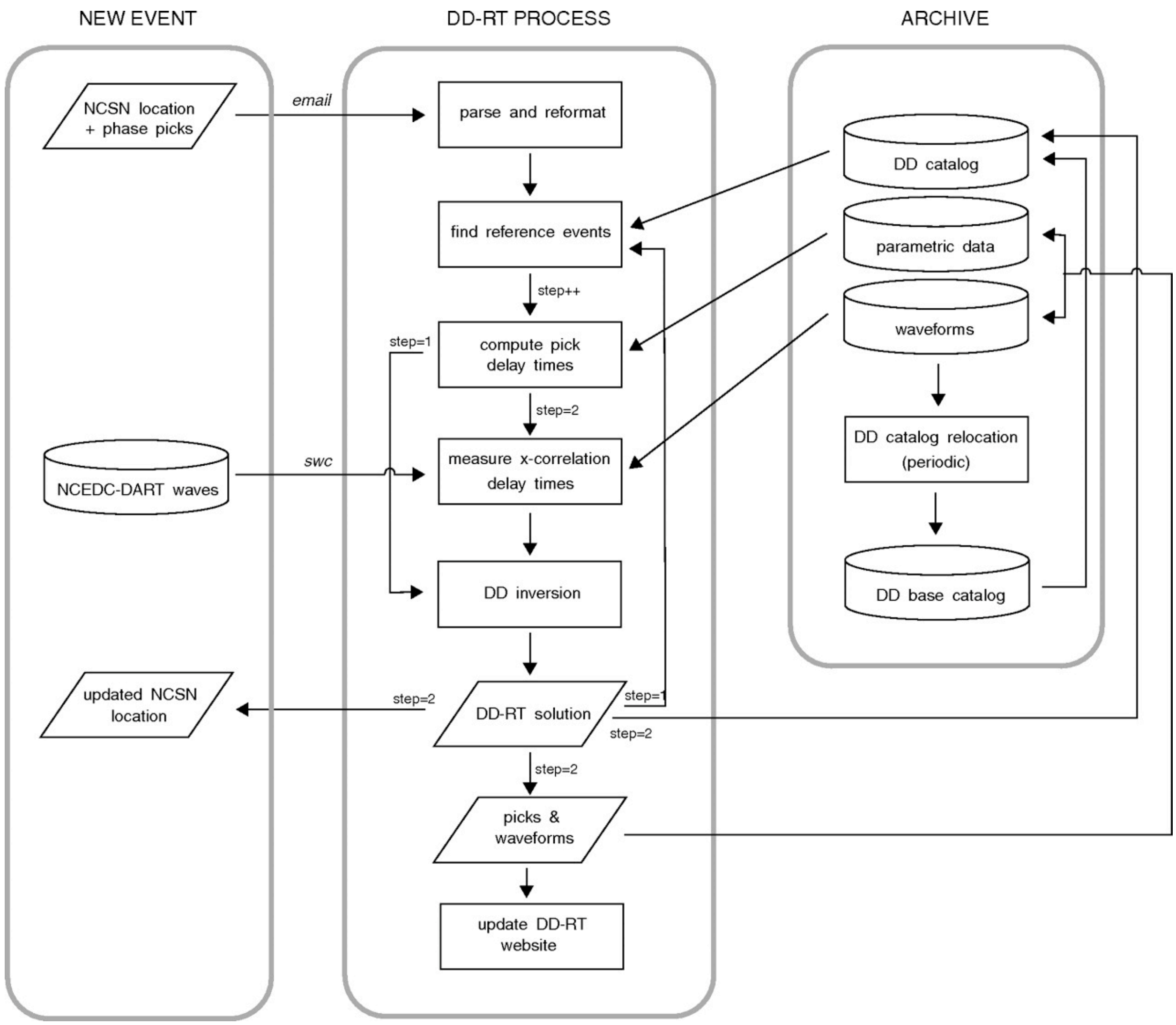

Figure 1 Detailed flow chart of the DD-RT procedure as implemented at LDEO for relocation of earthquakes in Northern California with real-time data feeds from the USGS in Menlo Park (NCSN) and UC Berkeley (NCEDC). *swc:* Simple Waveform Client (Doug Neuhauser, pers. comm., 2008). From Waldhauser (2009).

Data and Method

The DD-RT procedure relocates new events recorded at the NCSN network relative to the high-resolution double-difference (base) catalog described Waldhauser and Schaff (2008) by computing delay times (from phase picks and waveform cross-correlation) between a new event and its neighboring (reference) events in the base catalog and inverting these data using a real-time version, *hypoDD-RT*, of the double-difference algorithm *hypoDD* described in Waldhauser and Ellsworth (2000) and Waldhauser (2001). After a new event has been procesed at the NCSN, parametric data in Hypoinverse archive format (including phase picks, hypocenter location and origin time, and magntitude estimates) (Klein, 2002) are automatically being sent to the LDEO server via email and seismograms, two minutes long and starting 10 s before the origin time, are automatically requested from the NCEDC using the program *swc* (Simple Waveform Client; Doug Neuhauser, pers. communication; Romanowicz et al., 1994). The *swc* program access the Simple Wave Server (SWS) at the NCEDC which stores the real-time DART waveforms, and downloads the seismograms of new events in miniSEED format to the local disk. A search algorithm has been developed that assures a spatially homogeneous sub-sampling of the neighboring seismicity and the selection of optimal reference events (see Waldhauser, 2009). The parametric and waveform data for the reference events are retrieved from a locally stored copy of the entire NCSN archive from 1984-2003 (the period for which we currently have a DD base catalog), totaling about 700 Gb in size. The archive consists of ~7 million NCSN phase arrival time picks (mostly P-phases) and some 15 million waveforms in SAC format.

Differential arrival times between a new event and its reference events, measured at common stations from routine phase picks and/or from waveform cross-correlation, are simultaneously inverted in an iterative weighted least-squares procedure for adjustments in the location of the new event relative to its reference events in the DD catalog. We use a modified version of the double-difference algorithm *hypoDD* (Waldhauser and Ellsworth, 2000; Waldhauser, 2001) that uses only the equations that constrain a new event, *i*, relative to its *nref* reference events, *j*, in the DD catalog:

$$\frac{\partial t_k^{i=1}}{\partial \mathbf{m}} \Delta \mathbf{m}^i - \frac{\partial t_k^{j=1,nref}}{\partial \mathbf{m}} \Delta \mathbf{m}^j = dr_k^{ij} \tag{1}$$

where dr_k^{ij} is the residual between observed (dt^{obs}) and predicted (dt^{cal}) phase travel-time difference between a new event and its reference events observed at a common station, *k* (see also Eq. 18 in Waldhauser and Ellsworth, 2000). $\Delta\mathbf{m}$ are changes in the vector connecting their hypocenters through the partial derivatives of the travel times, *t*, for each event with respect to the vector, **m**, of the four unknowns.

The data is weighted according to an a priori quality weight and a dynamically adjusted weight that accounts for residual performance and inter-event distance during individual iterations (see Waldhauser and Ellsworth, 2000). Changes in the hypocentral parameters of the reference events, $\Delta\mathbf{m}^j$, are damped by adding to the system of double-difference equations four additional equations per event. The real-time DD algorithm has some similarity to the master event location method (e.g., Evernden, 1967), except that in the master event approach many slave events are relocated relative to one master event, while here we relocate one slave event (new event) relative to a set of master events (reference events). The DD-RT solutions are automatically being posted to a web site (www.ldeo.columbia.edu/~felixw/DDrtCISN) from which they can be openly accessed. The summary figure of an example event is shown in Figure 2.

We note that this procedure significantly improves the precision of a new event's location relative to the background seismicity. The accuracy of the absolute location depends on the location accuracy of the reference events in the DD base catalog. We refer to Waldhauser and Schaff (2008) for a detailed description and location error analysis of the DD base catalog for Northern California that is used in this study.

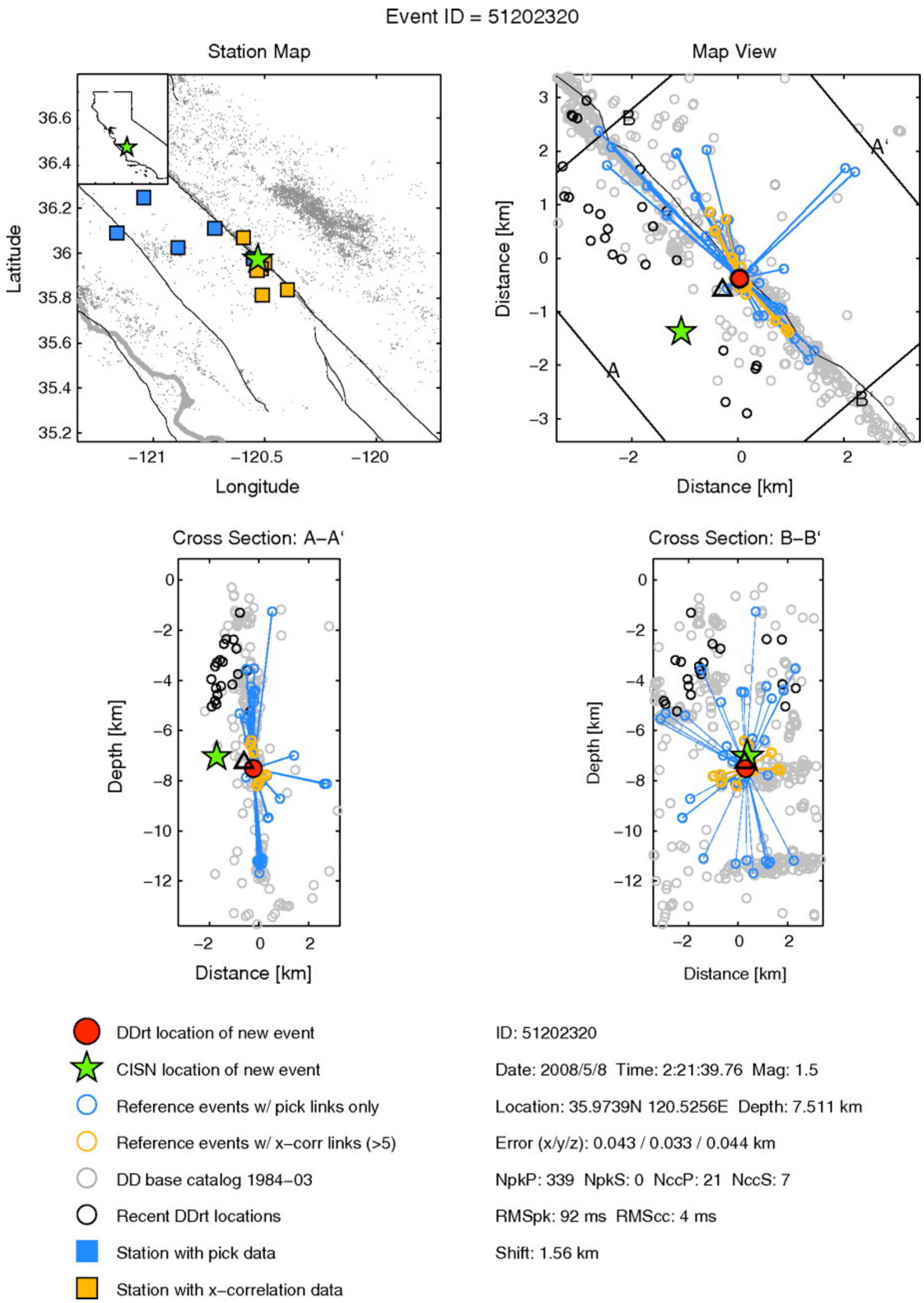

Figure 2 Example of automatically generated figure displaying and summarizing the DD-RT relocation performance and results for each event relocated. The event shown occurred on May 8, 2008 and has a magnitude of M_L=1.5. From Waldhauser (2009).

DD-RT results and performance

Pick based DD-RT solutions are computed within a few tens of seconds after receiving the initial location and phase information, using a single Sun UltraSparc IIIi processor. Final DD-RT solutions computed using cross-correlation data (if available) are typically available in less than a minute. New events that occur in areas of dense seismicity with numerous strongly correlated reference events may take longer, because of the additional time spent on accessing and cross-correlating the seismograms, and on inverting the larger system of DD equations. Real-time DD locations are obtained for all new events that have at least one reference event with at least four P-phase picks at common stations. New events that cannot be relocated, such as events that occur in areas with no historic activity or are recorded by new stations only, are flagged as such and added to the continuous DD catalog.

The DD-RT process automatically generates figures and tables that summarize key parameters describing the solution quality and characterize the hypocenter's location relative to the past seismicity. Figure 2 shows an example summary figure for an M_L=1.5 event that occurred on the San Andreas Fault near Parkfield on May 8, 2008. It displays the final DD-RT location (red dot in Figure 2) relative to its routinely determined solution (green star), and relative to the historic seismicity as present in the DD base catalog (gray dots). The example illustrates how the hypocenter of the event, initially located about 1.5 km to the south-west of the San Andreas fault, falls onto the main fault after relocation. The preliminary pick based DD solution for this event (shown as triangle in Figure 2) is less than 300 m from the final cross-correlation based DD location.

A back-testing scheme is employed to evaluate the overall performance of the DD-RT procedure in Northern California. We treat past events as 'new' events and relocate them relative to the DD base catalog while excluding data associated with the archived 'new' event. The DD-RT locations are then evaluated by their deviation from the corresponding location in the DD base catalog. We use a total of 2,360 events in the year 2002, selecting them randomly within cells of 20 x 20 km (max 50 events in each cell) to achieve uniform aerial coverage. The median RMS residual for all events after relocation is 4 msec for cross-correlation data and 30 msec for the pick data, down 94% and 52%, respectively, from the 64 msec seen in the routine locations. A summary of the relocation results showing differences between the two DD-RT locations (picks alone and picks combined with cross correlation data) and the corresponding locations in the DD base catalog is given in Figure 3 and Table 1. The median horizontal (vertical) deviation of the final DD-RT locations is 0.08 (0.24) km, compared to 0.34 (0.73) km in the routine locations. The highest precisions are achieved along the San Andreas fault and in Long Valley where epicentral differences have medians generally less than 0.05 km (Figure 3c, top panel). Differences in depths are generally less than 0.5 km, and less than 0.1 km in well monitored regions such as along the San Andreas fault and at the Geysers Geothermal Field (Figure 3c, bottom panel). Pick based DD-RT relocations are typically within 0.1 km horizontally and 0.5 km vertically along the San Andreas fault and Long Valley, and within 0.5 km horizontally and 1 km vertically elsewhere (Figure 3b).

The DD-RT locations of the 2,360 back-testing events in 2002 represent a significant improvement over the corresponding routine locations across most of Northern California. We achieve horizontal (vertical) location improvement of a factor of 10 or better for 12% (3%) of the seismically active area (as calculated by the number of cells). For a factor 5 or better these values are 41% (19%) and for a factor 2 or better 82% (57%). Epicenter locations improve in all areas, while depth locations appear to get worse in 3% of the area. The precision with which we can relocate new events relative to the background seismicity is increasing with time, as the density of earthquakes increases along active faults (Waldhauser and Schaff, 2008). This is especially true for events that occur in regions with low seismicity rates (Waldhauser, 2009).

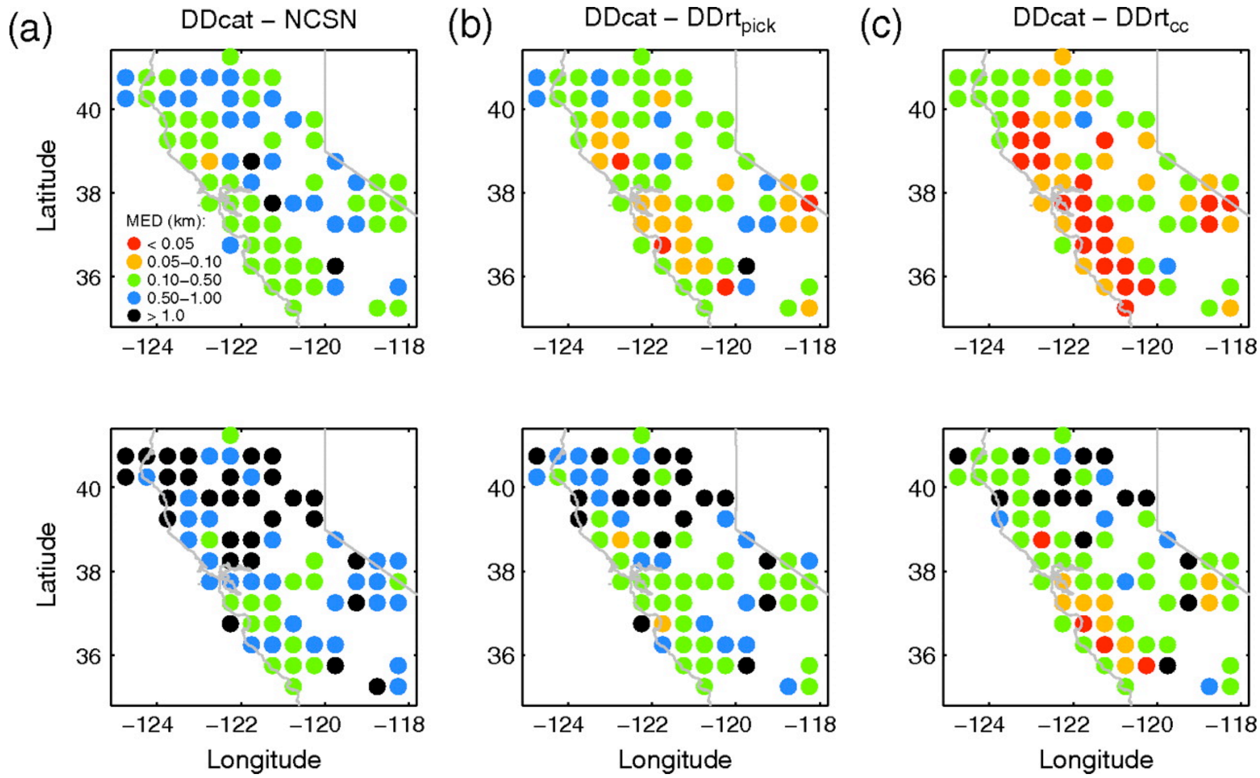

Figure 3 Results from back-testing the DD-RT procedure using 2360 events from 2002. Differences between epicenter (top panels) and depth (bottom panels) locations in the DD catalog and corresponding locations in the NCSN catalog (left), the pick-based DD-RT locations (middle panels), and the cross-correlation based DD-RT locations (right panels). Median differences are shown for events in 20 x 20 km cells.

Table 1 Back-testing results II: Deviations in horizontal and vertical directions of the NCSN and the pick and cross-correlation based DD-RT solutions from the corresponding locations in the DD base catalog for 2360 events that occurred across Northern CA in the year 2002.

	DX (km)			DY (km)			DZ (km)			RMS (msec)		
	med	mean	max	med	mean	max	med	mean	max	med	mean	max
NCSN	0.34	0.65	8.9	0.33	0.61	7.36	0.75	1.34	15.56	64	88	1797
DD-RT$_{pick}$	0.14	0.40	5.77	0.14	0.37	6.58	0.44	1.02	15.55	30	44	824
DD-RT$_{xcorr}$	0.08	0.29	5.73	0.07	0.27	9.00	0.24	0.80	15.55	4	10	661

2.2 Update of current (1984-2003) DD catalog

As mentioned in 'Overview of Investigations' we have experience a delay in receiving the waveform data necessary to update the existing DD catalog for Northern California (1984-2003; Waldhauser and Schaff, 2008) with more recent events. We have been able to acquire the data (2003-2008) from the NCEDC at the beginning of 2009, and have subsequently carried out the cross-correlation delay time measurements between the new events and the new and old events at common stations using the procedures described in Schaff and Waldahuser (2005). A total of 145,000 events have been added to the existing DD catalog of 311,000 events, and 0.5 billion cross-corelation delay times were added to the existing data base of 3 billion delay times. An initial double-difference inversion has been carried out and uncertainty estimates are currently being computed. Robust and resolution tests need to be carried out before it can be released to the scientific community.

We note that the catalog update was complicated by the fact that the CISN/NCSN undertook significant changes to parameters such as event IDs, station names, component names, and waveform format. These changes required modification to almost every piece of software and related databases used during the production of the existing DD catalog, including some 3 billion cross correlation delay times.

The updated DD catalog, when finalized, will serve as a new DD base catalog in the DD-RT system (see 'Final Remarks' below), which then will produce a continously updated DD catlaog for Northern California. The availability of continuously updated double-difference earthquake catalogs in near real-time is expected to have considerable social and economic impact in the evaluation and mitigation of seismic hazards, and the catalogs are particularly valuable to the research community as they provide fundamental data in the geophysical sciences. For example, rapid knowledge of precise aftershock locations may be useful to delineate the rupture area of the mainshock, which may help scientists assess the potential for and size of future aftershocks as well as provide critical source information for strong ground motion prediction. Changes in the recurrence intervals of repeating events can be monitored to infer changes in the loading rate, or the short-term evolution of stress concentrations can be tracked in near-real-time to better assess the potential occurrence of future events. Monitoring fine-scale changes in seismogenic behavior is particularly important along hazardous faults with well-characterized spatio-temporal behavior of past seismicity like the Hayward, Calaveras, and Central San Andreas faults.

4. References

Evernden, J.F. (1969). Identification of earthquakes and explosions by use of teleseismic data, *J. Geophys. Res.* **74,** 3828-3856.

Klein, F. W. (2002), User's Guide to HYPOINVERSE2000, a Fortran Program to Solve for Earthquake Locations and Magnitudes, *U. S. Geol. Surv. Open-File Rep*, *02-172*, 123pp.

Romanowicz, B., D. Neuhauser, B. Bogaert and D.Oppenheimer (1994), Accessing Northern California Earthquake Data Via Internet, EOS Trans A.G.U. **75**, 258-261.

Schaff, D.P. and F. Waldhauser (2005), Waveform cross-correlation-based differential travel-time measurements at the Northern California Seismic Network, *Bull. Seism. Soc. Am.*, *95*, 2446-2461.

Waldhauser, F. and W. L. Ellsworth (2000), A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, *Bull. Seism. Soc. Am.*, 90, 1353-1368.

Waldhauser, F. (2001), HypoDD: A computer program to compute double-difference hypocenter locations, *U.S. Geol. Surv. open-file report*, 01-113, Menlo Park, California.

Waldhauser, F. and D.P. Schaff, Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.*, 113, B08311, doi:10.1029/2007JB005479, 2008.

Waldhauser, F., Near-real-time double-difference event location using long-term seismic archives, with application to Northern California, *Bull. Seism. Soc. Am.*, 99, 2736-2848, doi:10.1785/0120080294, 2009.

5. Reports published related to this project

Peer-reviewed journal publications:

Waldhauser, F., Near-real-time double-difference event location using long-term seismic archives, with application to Northern California, *Bull. Seism. Soc. Am.*, 99, 2736-2848, doi:10.1785/0120080294, 2009.

Waldhauser, F. and D.P. Schaff, Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.*, 113, B08311, doi:10.1029/2007JB005479, 2008.

Conference abstracts:

Waldhauser, F. and D. P. Schaff, A cross-correlation double-difference catalog for northern California and its use in near real-time event relocation, 5th Annual Northern California Earthquake Hazards Workshop, Menlo Park, CA, January 23-24, 2008.

McCrory, P.A., J.L. Blair, F. Waldhauser, Mapping the Juan de Fuca slab beneath the Cascadia margin, GSA Annual Meeting, Portland, October 18-21, 2009.

Waldhauser, F., and D. Schaff, Near-real-time double-difference earthquake locations for Northern California, 6th Annual Northern California Earthquake Hazards Workshop, Menlo Park, CA, January 20-21, 2009.

Waldhauser F. and D. Schaff, Seismic monitoring using long-term archives: New directions and opportunities, Bi-Lateral Workshop Under the Sino-US Earthquake Studies Protocol,

Boulder, CO, November 11-14, 2008.
Waldhauser, F, A real-time procedure for double-difference event location: performance evaluation in Northern California, IRIS Workshop, Stevenson, WA, June 4-6, 2008.
Simpson, R., D. Oppenheimer, W. Bakun, R. Graymer, F. Waldhauser, and D. Schaff, A new look at Calaveras Fault zonation using double-difference relocated earthquakes, Third Conference on Earthquake Hazards in the Eastern San Francisco Bay Area, CSU East Bay, October 22-26, 2008.
Waldhauser, F. and D. P. Schaff, A cross-correlation double-difference catalog for northern California and its use in near real-time event relocation, 5th Annual Northern California Earthquake Hazards Workshop, Menlo Park, CA, January 23-24, 2008.

6. Available Data and Products

The near-real-time double-difference hypocenter solutions for earthquakes in Northern California produced by the DD-RT method developed under this grant are continuously being made available at www.ldeo.columbia.edu/~felixw/DDrtCISN on a beta-test basis. The current DD catalog (NCAeq_DD.1983-2003.v1.0) can be downloaded via a link from this same website, as will future updates of the catalog.

Contact: Dr. Felix Waldhauser
Lamont-Doherty Earth Observatory
61 Route 9W
Palisades, NY 10964
845 365 8538
felixw@ldeo.columbia.edu

7. Final Remarks

We are close to completing the work of updating the existing DD catalog to include all events until the end of 2008. We will then be replacing the current (2003 edition) DD base catalog in the DD-RT system with the new, updated DD catalog. An additional update to include the 2009 events is planned for this Summer. The 2009 results are then compared to the DD-RT solution that were automatically computed during 2009, which gives us the necessary information to fine tune the DD-RT parameters. This work is being supported by a current USGS-NEHRP grant, G09AP00018.

Under a separate NEHRP grant, G10AC00040, we, in collaboration with P. Friberg (ISTI), will be installing an adapted DD-RT software package on a dedicated workstation at the USGS in Menlo Park. The DD-RT process will be integrated into the real-time system at the California Integrated Seismic Network (CISN) to allow a more seamless interaction with the existing databases and archiving systems.